作業3: 決策問題決定一切

**Q1. 典型的資料分析可以用以解決下面六種問題。**

Data analysts typically work with six problem types

1. Making predictions 2. Categorizing things 3. Spotting something unusual 4. Identifying themes 5. Discovering connections 6. Finding patterns

### **Making predictions**

Using data to make informed decisions about how things may be in the future.

A company that wants to know the best advertising method to bring in new customers. Analysts with data on location, type of media, and number of new customers acquired as a result of past ads can't guarantee future results, but they can help predict the best placement of advertising to reach the target audience.

### **Categorizing things 分類→瞭解這一類的特徵是什麼**

Grouping data based on common features.

An example of a problem requiring analysts to categorize things is a company's goal to improve customer satisfaction. Analysts might classify customer service calls based on certain keywords or scores. This could help identify top-performing customer service representatives or help correlate certain actions taken with higher customer satisfaction scores.

### **Spotting something unusual**

Identifying data that is different from the norm.

A company that sells smart watches that help people monitor their health would be interested in designing their software to spot something unusual. Analysts who have analyzed aggregated health data can help product developers determine the right algorithms to spot and set off alarms when certain data doesn't trend normally.

### **Identifying themes →上下文、內文分析，找到上下文關聯性，要先做分類再找類別跟類別之間的關聯性。**

Recognizing broader concepts and trends from categorized data.

User experience (UX) designers might rely on analysts to analyze user interaction data. Similar to problems that require analysts to categorize things, usability improvement projects might require help prioritize the right product features for improvement. Themes are most often used to help researchers explore certain aspects of data. In a user study, user beliefs, practices, and needs are examples of themes.

By now you might be wondering if there is a difference between categorizing things and identifying themes. The best way to think about it is: categorizing things involves assigning items to categories; identifying themes takes those categories a step further by grouping them into broader themes.

### **Discovering connections → 找到因果關係**

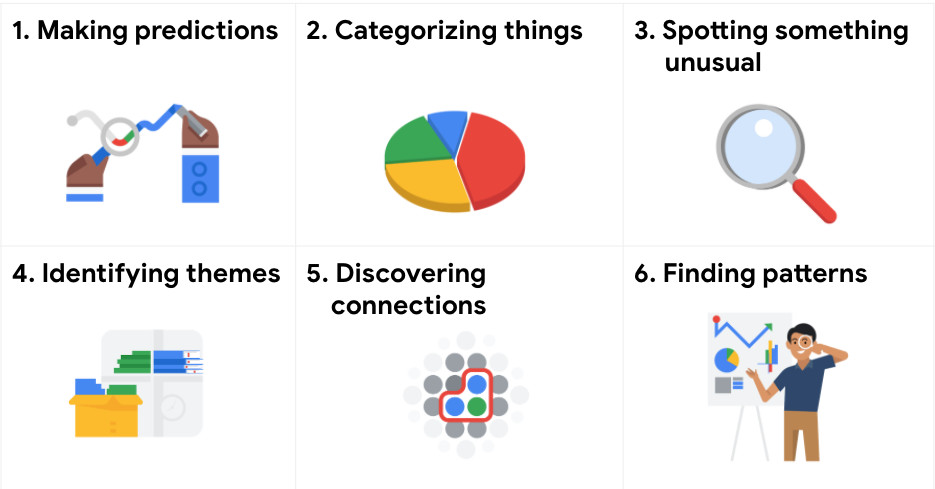
Identifying similar challenges across different entities—and using data and insights to find common solutions.

A third-party logistics company working with another company to get shipments delivered to customers on time is a problem requiring analysts to discover connections. By analyzing the wait times at shipping hubs, analysts can determine the appropriate schedule changes to increase the number of on-time deliveries.

### **Finding patterns → 找到關聯性**

Using historical data about what happened in the past to understand how likely it is to happen again.

Minimizing downtime caused by machine failure is an example of a problem requiring analysts to find patterns in data. For example, by analyzing maintenance data, they might discover that most failures happen if regular maintenance is delayed by more than a 15-day window.



接下來，請依照您抽到的牌卡順序，各給出例子，並應用可能的open data**資料說明其輸入和輸出資料**。

例如:

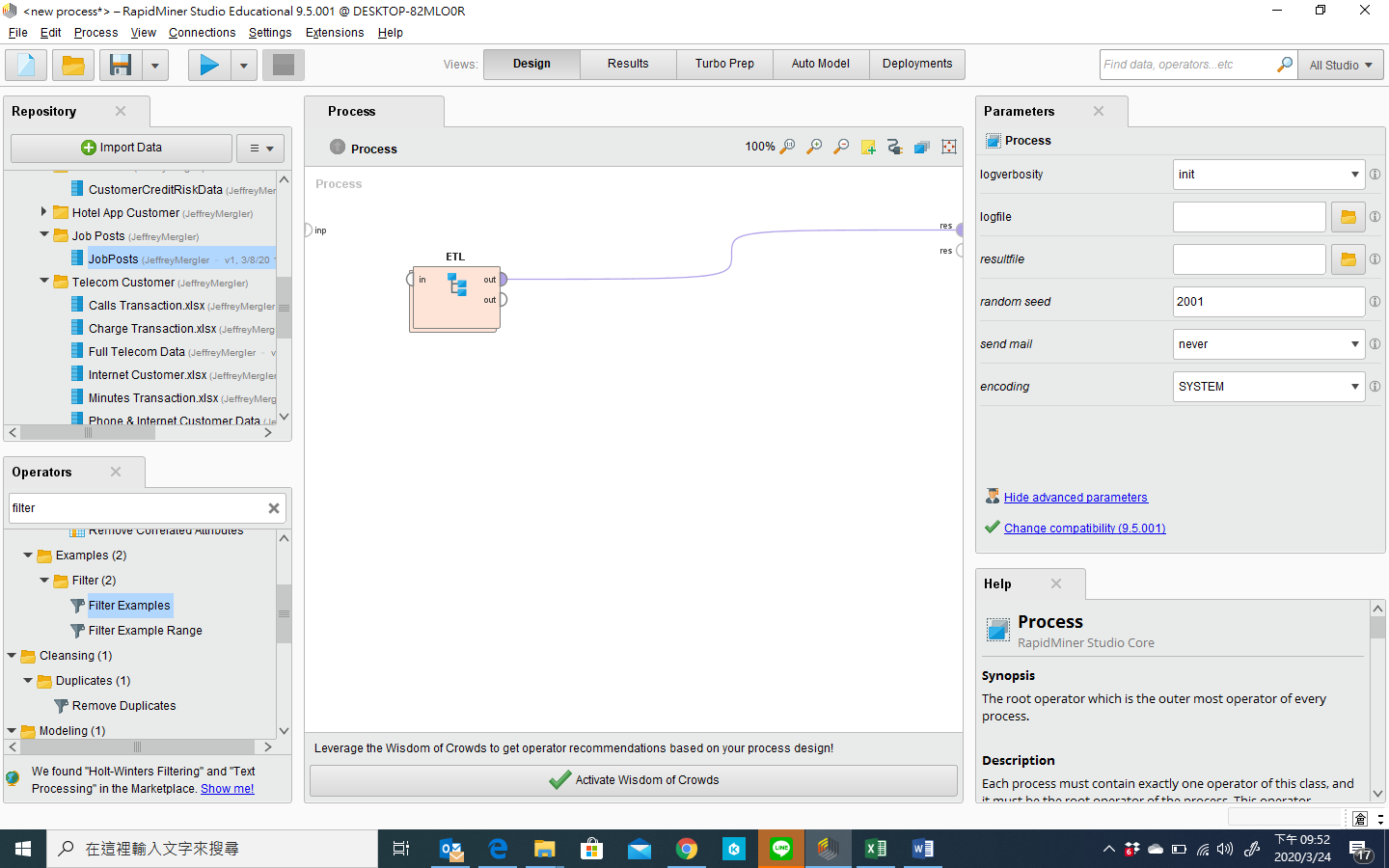
1. Making predictions（行）：Ex. 預測房價  
   Input: 高速公路發布路段即時路況資料

Output: 預測未來交通堵塞狀況

1. Categorizing things （老）：Ex. 骨架判斷-長輩是否跌倒；或銀行是否遭到打劫
2. Spotting something unusual （病）：
3. Identifying themes （衣）：
4. Discovering connections （育）
5. **Finding patterns（樂）：**

**Q2 薪資大不同**

2-1 讀入檔案(salary.csv),請大概說明一下您對於資料的了解(5%),並進行「必要」之資料前處理,並請將相關前處理存於ETL子流程如下圖(請說明您做了什麼,為什麼) (10%)



2-2 .請試著解決並以salary為預測標的，貼出您的決策樹,並加以說它的意涵(10%)

2-3 請列出您的混肴矩陣,並請加以說明其意涵，除此之外，我還希望您能列出您模型的下述４個指標，並請簡略說明他們的計算方式及含意(5%)



2-4最後，請說明您的管理意涵？ (10%) → 這題最難寫

請將您的process存檔為學號-1, 如: 106AB001\_1.rmp檔